(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 January 2005 (27.01.2005)

PCT

(10) International Publication Number
WO 2005/008571 A1

(51) International Patent Classification[7]: G06K 9/62

(21) International Application Number: PCT/US2004/021511

(22) International Filing Date: 2 July 2004 (02.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/615,885 8 July 2003 (08.07.2003) US

(71) Applicant (*for all designated States except US*): **COMPUTER ASSOCIATES THINK, INC.** [US/US]; One Corporate Associates Plaza, Islandia, NY 11749 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **DUAN, Baofu** [CN/US]; 2235 Overlook Road, Apartment 208, Cleveland Heights, OH 44106 (US). **MENG, Zhuo** [CN/US]; 8374 Chestnut Boulevard, Broadview Heights, OH 44147 (US). **PAO, Yoh-Han** [US/US]; 2721 Scarborough Road, Cleveland Heights, OH 44106 (US).

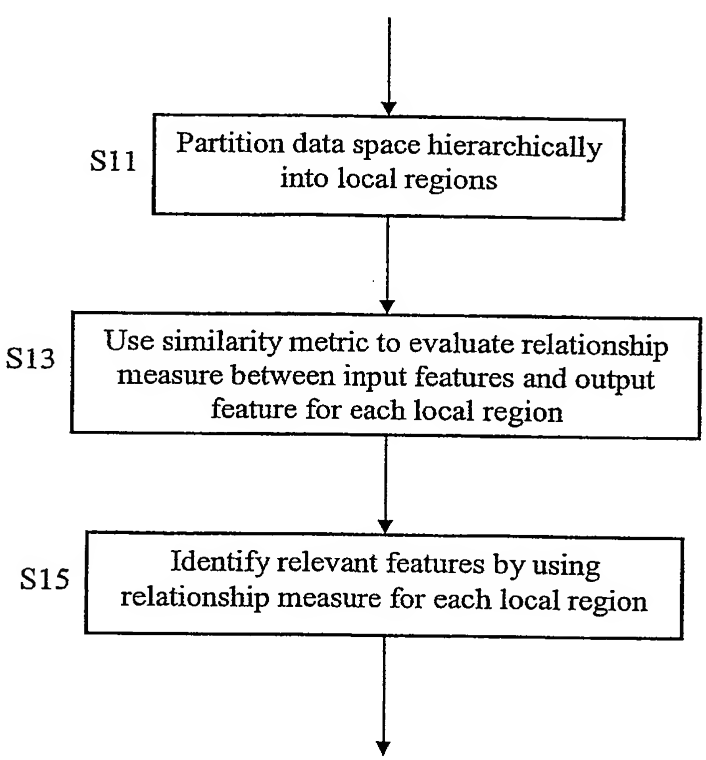
(74) Agents: **JAWORSKI, Richard, F.** et al.; Cooper & Dunham LLP, 1185 Avenue of the America, New York, NY 10036 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,

[Continued on next page]

(54) Title: HIERARCHICAL DETERMINATION OF FEATURE RELEVANCY



(57) Abstract: Methods for feature selection based on hierarchical local-region analysis of feature relationships in a data set are provided.

WO 2005/008571 A1

SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

HIERARCHICAL DETERMINATION OF FEATURE RELEVANCY

TECHNICAL FIELD

This application relates to pattern recognition and data mining. In particular, the application relates to feature analysis for pattern recognition and data mining.

DESCRIPTION OF RELATED ART

Feature selection is of theoretical interest and practical importance in the practice of pattern recognition and data mining. Data objects typically can be described in terms of a number of feature values. The task is to determine what feature or subset of features is to be used as the basis for decision making in classification and for other related data mining tasks. Although objects or data entities can be described in terms of many features, some features may be redundant or irrelevant for specific tasks, and therefore instead may serve primarily as a source of confusion. It is not necessarily true that a larger number of features provides better results in task performance. Inclusion of irrelevant features increases noise and computational complexity. In addition, for any one specific task, different subsets of features might be relevant in different regions of input data space. Therefore, feature selection is a matter of considerable interest and importance in multivariate data analysis.

For example, when a specific behavior or output of a specific system is modeled, it is typically desirable to include only parameters that contribute to the modeled system behavior and not other parameters which contribute to other behaviors of the system but are not particularly relevant to the specific modeled behavior.

In a classification task, a process for identifying relevant features can usually be formalized to specify a criterion for class assignment followed by an evaluation of the ability of the specified criterion to serve as a basis for class separation or for minimizing the degree of overlap between different classes. Features can then be evaluated on a basis of how effective they are when used in combination with the specified criterion.

As a slight variation to the process described above, instead of selecting a set of features for a specific criterion, one can rank the features that contribute to separation of classes. One issue that is often presented is how to search an optimum group of features for a specific criterion, where the number of possible groups of features is combinatorial. Many methods have been proposed involving or based on neural networks, genetic algorithms, fuzzy sets, or hybrids of those methodologies.

However, there is a need for improved methods for feature selection.

SUMMARY

The application provides a method for feature selection based on hierarchical local-region analysis of feature relationships in a data set. In one embodiment, the method includes partitioning hierarchically a data space associated with a data set into a plurality of local regions, using a similarity metric to evaluate for each local region a relationship measure between input features and a selected output feature, and identifying one or more relevant features, by using the similarity metric for each local region.

According to another embodiment, a method for feature selection based on hierarchical local-region analysis of feature characteristics in a data set, includes partitioning hierarchically a data space corresponding to a data set into a plurality of local regions, using a relationship measure to evaluate for each local region a correlation between input feature values on the one hand and a selected output on the other hand, and determining a relevancy of a selected feature by performing a weighted sum of the relationship measure for the feature over the plurality of local regions.

Hierarchical local-region analysis is the key to successful identification of relevant features. As it is

evident in examples provided below, neither too few nor too many local regions would yield satisfactory results.

BRIEF DESCRIPTION OF THE DRAWINGS

The features of the present application can be more readily understood from the following detailed description with reference to the accompanying drawings wherein:

FIG. 1 shows a flow chart of a method, according to one embodiment, for feature selection based on hierarchical local-region analysis of feature characteristics in a data set;

FIG. 2 shows a flow chart of a method for feature selection based on hierarchical local-region analysis of feature characteristics in a data set, according to an alternative embodiment of the present application;

FIG. 3 shows a flow chart of an exemplary embodiment of a method for hierarchical determination of feature relevancy;

FIG. 4 shows a three-dimensional plot of an extended parity-2 problem;

FIG. 5 shows a plot which demonstrates feature relevancies at different levels for the extended parity-2 problem;

FIG. 6 shows performance of neural net modeling without and with noise features; and

FIG. 7 shows a plot which demonstrates feature relevancies at different levels for the extended parity-5 problem.

DETAILED DESCRIPTION

This application provides tools (in the form of methodologies and systems) for identifying relevant features (from a set of available or specified features), for example, through feature ranking and/or selection, for feature analysis. The tools may be embodied in one or more computer programs stored on a computer readable medium and/or transmitted via a computer network or other transmission medium.

Methods for feature selection based on hierarchical local-region analysis of feature characteristics in a data set are described in this application. A method for feature selection , according to one embodiment, will be described with reference to FIG. 1. A data space associated with a data set is partitioned hierarchically into a plurality of local regions (step S11). A similarity metric is used to evaluate for each local region a relationship measure between input features and a selected output feature (step S13). One or more relevant features is identified by using the relationship measure for each local region (step S15). The method may further include

determining a feature relevancy of a selected feature by performing a weighted sum of the relationship measures for the selected feature over the plurality of local regions. The weights for the weighted sum may be based on sizes of the respective local regions.

The partitioning of the data space into the plurality of local regions can be performed by hierarchical clustering of the data set in a plurality of levels. Feature relevancies can be determined for each of the input features based on the relationship measure at each level of the hierarchical clustering, and the relevant features identified based on the feature relevancies.

The method may further include determining for each local region a corresponding subset of relevant features based on the relationship measure for the local region. The subsets of relevant features for respective local regions may be non-identical. The local regions may be nonoverlapping.

The similarity metric may be linear, and may include a projection or distance. The relationship measure may include a correlation or R^2.

A method for feature selection based on hierarchical local-region analysis of feature characteristics in a data set, according to another embodiment, will be explained with reference to FIG. 2. A data space corresponding to a

data set is partitioned hierarchically into a plurality of local regions (step S21). A similarity metric is used to evaluate for each local region a relationship measure between input feature values on the one hand and a selected output on the other hand (step S23). A relevancy of a selected feature is determined by performing a weighted sum of the relationship measures for the feature over the plurality of local regions (step S25). The weights for the weighted sum may be based on sizes of the respective local regions. The method may further comprise ranking the input features according to the corresponding feature relevancies of the input features. The local regions may be nonoverlapping.

The partitioning of the data space may be performed through hierarchical clustering of the data set in a plurality of cluster levels. The method may further include identifying relevant features at each level of the hierarchical clustering and determining corresponding feature relevancies.

Feature analysis can be motivated by the need to pick the most relevant features from all of the available ones, given a specific dependent feature or quality. This disclosure describes hierarchical determination of feature relevancy (HDFR) which can be applied to feature selection and/or ranking on the basis of relevancy to a task at hand.

For an example of modeling a specific behavior, or output, of a specific system, the selection criterion can be the relevancy of a feature to the specific behavior output. In order to assess relevancy of a feature, one can simply compute the correlation between the feature and the specific behavior output. If a strong correlation exists, the feature is apparently relevant to the specific output. However, although a feature may not show strong correlation over the whole range of data input values, it might nevertheless show strong correlation over different ranges. Such a feature can still be considered relevant and thus selected.

Hierarchical determination of feature relevancy can be used for the task of feature selection based on hierarchical local-region analysis of feature characteristics. Hierarchical clustering may be combined with various linear or nonlinear similarity metrics. In any event, hierarchical clustering can be used to delineate the partitioning of the entire body of input data into non-overlapping local regions.

In each local region, there might be a corresponding subset of features that is relevant according to the metric being used for the task in question. Different regions of input data space may or may not have the same subset of features. In other words, a feature or subset of features

might not show strong relevancy to a particular task over the entire range of data but might show strong relevancy over different delineated local regions. Such a feature can still be considered relevant and can be identified for use in the appropriate regions. Region delineation enhances a likelihood that the subsequent feature selection process successfully identifies the relevancies of features for a particular local region.

According to one embodiment in which HDFR is applied to system modeling, hierarchical clustering can be used to partition data space into local regions and a similarity metric is used to evaluate relationship measures between input feature values and system output for entities in each local region. The weighted sum of the relationship measures for a selected feature evaluated over all of the local regions can be used as a measure of the relevancy of the selected feature for a selected task. By applying this technique to a set of features, a subset of relevant features can be identified. For other circumstances, feature relevancy might be evaluated on the basis of maximum similarity. In addition, different subsets of relevant features can be identified for different regions of input data space.

The relevancy data structures can be managed through hierarchical clustering. The relevancies of features in

local regions at one level of the hierarchy can be considered together to determine the relevant features for that level. The relevant features for the problem at large can be derived from a consideration of the evaluations over the local regions at each level of the hierarchy. The hierarchical approach increases a probability of discovering subtle relevancies by avoiding accidental cancellation of correlation and also helps to prune accidental relationships.

For illustration purposes, additional exemplary embodiments are described below.

An exemplary embodiment of hierarchical determination of feature relevancy which utilizes a linear metric is described below. This exemplary embodiment may be applied to discover feature relevancies of numeric data with the assumption that the input features have a certain numeric relationship with the output. Hierarchical clustering is used to partition and transform data into groups of points in hyper-spherical local regions. A linear metric (for example, R-squared) is used to evaluate the relationship between input features and the output. R-squared values over all of the local regions are summarized as the relevancies of input features.

The embodiment can be analogized to an example of approximating scalar function defined in n-dimensional

space. Given a function $y = f(X)$, where $X = (x_1, x_2, \ldots, x_n)^T$ is the n-dimensional input variable and y is the output scalar variable, if the function $f()$ is differentiable at point X_0, (i.e., the first partial derivative functions $f^{(1)}(X) = (\partial f/\partial x_1(X), \partial f/\partial x_2(X), \ldots, \partial f/\partial x_n$ exists), then a tangent function $L(X) = f(X_0) + f^{(1)}(X_0)\ (X-X_0)$ is the linear approximation of $f(X)$ in the neighbor region of X_0. The approximation error can be as small as desired if the neighbor region is small enough. For a particular system, the piecewise linear approximation method partitions the system data space into many small regions and builds a linear approximation model in each local region. Each localized linear approximation model is valid only in its corresponding local region and the linear models together serve as a linear approximation model for the system.

An exemplary embodiment of hierarchical determination of feature relevancy which adapts the piecewise linear approximation technique, rather than building a very accurate linear approximation for the problem, can evaluate the correlations between input features and the output feature in each of the local regions based on the assumption that the system can be linearly approximated in the local regions. After the correlations are evaluated, a linear metric can be used to evaluate the similarity between input feature values and the system output for

entities in each local region.

A hierarchical clustering technique can be used to partition a data space into local regions. One embodiment is explained with reference to FIG. 3. The data space is partitioned initially into two regions (step S31). For each of the regions in the present level of the hierarchy, feature relevancies are evaluated based on samples in the region (step S32). The feature relevancy of a feature can be measured by the R-squared value between the input feature and the output. Feature relevancies in two local regions are weighted based on the size of the local regions and then summed together (i.e. a weighted sum) as the feature relevancies in the present level (step S33). The feature relevancies in the level are used to identify relevant features which have significantly larger relevancies than the others (step S34). If no new relevant features can be identified for a certain number of levels (step S35, "NO") or a specified maximum number of levels is reached (step S36, "YES"), the feature relevancies can be summarized at all of the levels and a list of relevant features and their relevancies provided (step S37). The local regions in the current level are split further for the next level (step S31), until no new relevant features can be identified for a specified or predetermined number of iterations or a specified maximum number of levels is

reached.

The performance of hierarchical determination of feature relevancy is examined and explained below with two examples. One example is the extended parity-2 problem and the other is the extended parity-5 problem. The extended parity-2 and parity-5 problems are derived from the well-known parity-2 and parity-5 problems, but extended to use inputs and output of continuous values. Some random noise inputs are also added for determining whether HDFR can identify the relevant inputs from the noise inputs. The extended parity-5 problem is a more complex task and is used for comparison with the extended parity-2 problem.

The parity-2 problem is a well-known problem. In this problem, the output is the mod-2 sum of two binary input features. The partity-2 problem is extended by using continuous inputs and output. The following nonlinear equation can be used to simulate the problem:

$$y = x_1 + x_2 - 2*x_1*x_2$$

where x_1, x_2 and $y \in [0, 1]$.

A 3-D plot of the above equation is shown in FIG. 4. For testing purpose, 8 random input features, x_3 to x_{10}, are added as noise and 500 samples are randomly generated. The task is to identify the relevant features, x_1 and x_2, from the noise features, x_3 to x_{10}.

HDFR was used to partition the extended parity-2 data

space into as many levels as possible and evaluate the relevancy values of the input features at each level. FIG. 5 shows how the feature relevancies vary at different levels. In level 0 (i.e., the original data space), x_1 and x_2 are not significantly different from other noise features x_3 to x_{10}. In level 1, x_1 is identifies as a relevant feature. In level 2 (or further), both x_1 and x_2 are identified as relevant features. One interesting thing is that in level 10 and beyond, the relevancies of x_1 and x_2 are again not significantly different from other noise features x_3 to x_{10}. This is because of the limited number of samples. When the level goes higher, the number of samples in each local region becomes smaller. When the number of samples in a region is too small, the collection of samples in the region does not contain enough information to differentiate the relevant features from the noise features.

With use of neural net modeling technology, one might hypothesize that it is possible to feed all of the data to a neural net and see whether the model yields any sensible result. However, such practice is likely to yield disappointing results (even though neural net generally is an effective modeling tool). As with any modeling technique, one frequently faces the problem of "the curse of dimensionality." This problem, stated simply, is that an

exponential increase of the number of observations is needed in order to achieve the same level of detail for adding extra number of features. While neural nets may be better at coping with higher dimensions, trimming out irrelevant features typically yields much better results than adding more observations.

Two neural net models, one with all of the 10 input features (i.e. including the noise features) and the other with only the 2 relevant input features (i.e. x_1 and x_2), were utilized to demonstrate that use of only relevant features improves the quality of modeling. For comparison, two learning technique are used to build the neural net models, one being the traditional backpropagation (BP) learning technique using one hidden layer and three hidden nodes in the hidden layer net. The other uses radial basis functions net. FIG. 6 presents the results of the modeling. The values of four performance parameters are shown in FIG. 6, including the time expended to train the model (in seconds), degree of freedom (DOF) [which measures the complexity of the neural net model], mean squared error (MSE) for the training data set and ANOVA R-squared which measures how well the prediction of the neural net model matches the true output. The results show that the neural net models trained with the 2 relevant input features are superior to the neural net models trained with the 10 input

features in all of the four performance parameters.

Similar to the parity-2 problem but much more complex, the parity-5 problem has five input features. The output is the mod-2 sum of the five input features. The parity-5 problem also is extended by using continuous inputs and output. The five input features are x_1 to x_5. Also 5 random noise features, x_6 to x_{10}, are added and 1000 samples are randomly generated. The task is to identify the relevant features, x_1 to x_5, from the noise features, x_6 to x_{10}.

FIG. 7 shows the feature relevancies values at different levels. As can be seen in FIG. 7, the extended parity-5 problem is actually more complex than the extended parity-2 problem. Only x_3 and x_5 can be selected out in level 2. The process further selects x_2 in level 4 and x_4 in level 8. It is noted that x_1 is not selected out until level 10. Noise features x_6 to x_{10} are identified as irrelevant features. In level 12 and beyond, the relevancies of x_1 to x_5 are not significantly different from noise features x_6 to x_{10}.

This disclosure describes hierarchical determination of feature relevancy, which can be used to solve the task of feature selection based on hierarchical local-region analysis of feature characteristics. Hierarchical determination of feature relevancy is straightforward and much more efficient as compared with feature selection

techniques based on optimization search. HDFR is also very effective due to the hierarchical local region delineation. In addition, HDFR is scalable to handle a very large number of input features.

Some examples are discussed herein to show that HDFR is very effective for identifying relevant features which have subtle nonlinear relationship to the output even though the input features may not be correlated to the output in the whole data range. Although the exemplary embodiments of hierarchical determination of feature relevancy presented in this disclosure are adapted for determining feature relevancies for problems with numeric relationship, other implementations of HDFR can follow a similar process to solve problems with complex relationship, such as categorical and rule-based relationship. In such cases, the appropriate region delineation methods and similarity metrics can be used with HDFR.

Hierarchical determination of feature relevancy can be used to identify relevant features for a specific outcome. For example, HDFR can be applied in process (or system) monitoring, such as to identify relevant features which would trigger a need for adjustments to setpoints of the process or system, for example, when (or ideally before) a problem arises in the process or system, or adjustments

would facilitate a desired process output. For the exemplary case of modeling a system, the user can create a leaner and better performing model of a system by removing irrelevant features.

In addition, HDFR can be applied to a data set of historical samples of viral behavior in an information technology (IT) system to extract relevant features. The extracted features can be the basis for rules added to a rule-based security monitor which would, for example, trigger a security alert if the features are detected in the system when the monitor is deployed on-line.

As another example, HDFR can be applied to a consumer profile data set to extract relevant features from patterns in the data set which are associated with specific buying tendencies, or historical stock market data to determine relevant features in a bull market or bear market.

The exemplary embodiments described above are illustrative, and many variations can be introduced on these embodiments without departing from the spirit of the disclosure or from the scope of the appended claims. For example, elements and/or features of different exemplary embodiments may be combined with each other and/or substituted for each other within the scope of this disclosure and appended claims.

As another example, an alternative technique other

than hierarchical clustering may be used to generate the hierarchical partition of regions. In addition, other relevancy metrics may be used instead of R^2.

This application claims the priority of U.S. application Serial No. 10/615,885, filed July 8, 2003 and entitled "HIERARCHICAL DETERMINATION OF FEATURE RELEVANCY", which is incorporated herein in its entirety by reference.

What is claimed is:

1. A method for feature selection based on hierarchical local-region analysis of feature characteristics in a data set, comprising:

partitioning a data space associated with a data set into a hierarchy of pluralities of local regions;

using a similarity metric to evaluate for each local region a relationship measure between input features and a selected output feature; and

identifying one or more relevant features, by using the relationship measure for each local region.

2. The method of claim 1 further comprising:

determining a feature relevancy of a selected feature by performing a weighted sum of the relationship measures for the selected feature over the plurality of local regions.

3. The method of claim 2, wherein weights for the weighted sum are based on sizes of the respective local regions.

4. The method of claim 1, wherein the partitioning of the data space into the hierarchy of pluralities of local regions is performed by hierarchical clustering of the data

set in a plurality of levels.

5. The method of claim 4, wherein feature relevancies are determined for each of the input features based on the relationship measures at each level of the hierarchical clustering and the relevant features are identified based on the feature relevancies.

6. The method of claim 1 further comprising:

determining for each local region a corresponding subset of relevant features based on the relationship measures for the local region.

7. The method of claim 6, wherein the subsets of relevant features for respective local regions are non-identical.

8. The method of claim 1, wherein the local regions are nonoverlapping.

9. The method of claim 1, wherein the similarity metric is linear.

10. The method of claim 1, wherein the similarity metric includes a projection or distance.

11. The method of claim 1, wherein the relationship measure includes a correlation.

12. The method of claim 1, wherein the relationship measure includes R^2.

13. A computer system, comprising:

a processor; and

a program storage device readable by the computer system, tangibly embodying a program of instructions executable by the processor to perform the method claimed in claim 1.

14. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform the method claimed in claim 1.

15. A computer data signal transmitted in one or more segments in a transmission medium which embodies instructions executable by a computer to perform the method claimed in claim 1.

16. A method for feature selection based on hierarchical local-region analysis of feature

characteristics in a data set, comprising:

partitioning a data space corresponding to a data set into a hierarchy of pluralities of local regions;

on each level of the hierarchy, using a similarity metric to evaluate for each local region in the level a relationship measure between input feature values on the one hand and a selected output on the other hand; and

determining a relevancy of a selected feature by performing a weighted sum of the relationship measures for the feature over the plurality of local regions at appropriate levels.

17. The method of claim 16, wherein the partitioning of the data space is performed through hierarchical clustering of the data set in a plurality of cluster levels.

18. The method of claim 17 further comprising:

identifying relevant features at each level of the hierarchical clustering and determining corresponding feature relevancies.

19. The method of claim 16, wherein weights for the weighted sum are based on sizes of the respective local regions.

20. The method of claim 16 further comprising:

ranking the input features according to the corresponding feature relevancies of the input features.

21. The method of claim 16, wherein the local regions are nonoverlapping.

22. The method of claim 16, wherein the similarity metric is linear.

23. The method of claim 16, wherein the similarity metric includes a projection or distance.

24. The method of claim 16, wherein the relationship measure includes a correlation.

25. The method of claim 16, wherein the relationship measure includes R^2.

26. A computer system, comprising:

a processor; and

a program storage device readable by the computer system, tangibly embodying a program of instructions executable by the processor to perform the method claimed

in claim 16.

27. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform the method claimed in claim 16.

28. A computer data signal transmitted in one or more segments in a transmission medium which embodies instructions executable by a computer to perform the method claimed in claim 16.

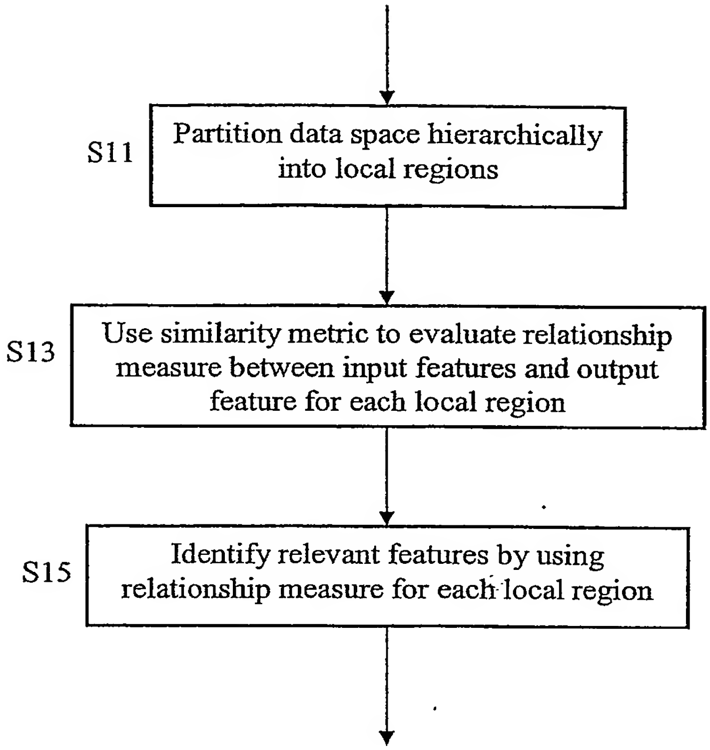

S11
Partition data space hierarchically into local regions
S13
Use similarity metric to evaluate relationship measure between input features and output feature for each local region
S15
Identify relevant features by using relationship measure for each local region

FIG. 1

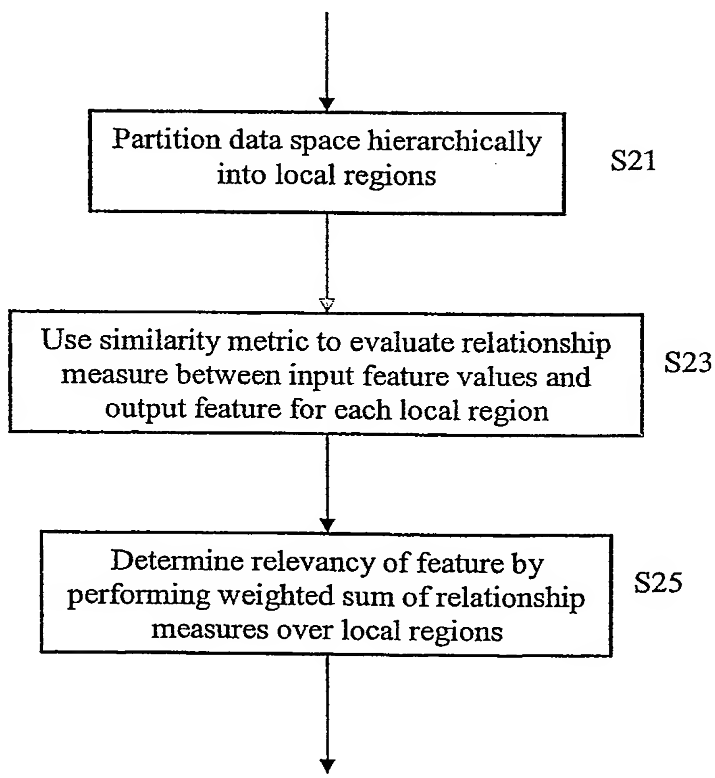

Partition data space hierarchically into local regions
S21
Use similarity metric to evaluate relationship measure between input feature values and output feature for each local region
S23
Determine relevancy of feature by performing weighted sum of relationship measures over local regions
S25

FIG. 2

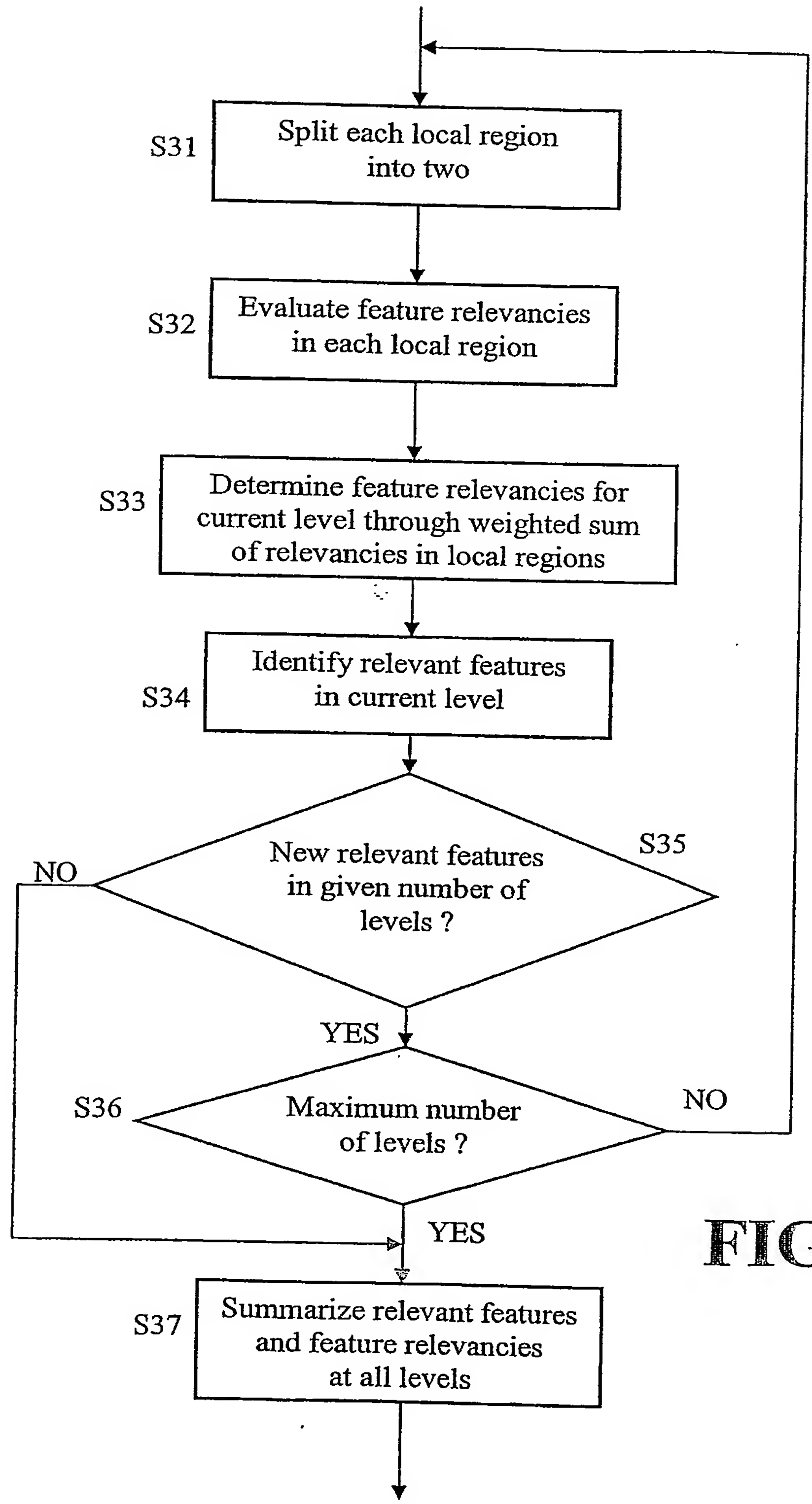

S31
Split each local region into two
S32
Evaluate feature relevancies in each local region
S33
Determine feature relevancies for current level through weighted sum of relevancies in local regions
S34
Identify relevant features in current level
S35
New relevant features in given number of levels ?
NO
YES
S36
Maximum number of levels ?
NO
YES
S37
Summarize relevant features and feature relevancies at all levels

FIG. 3

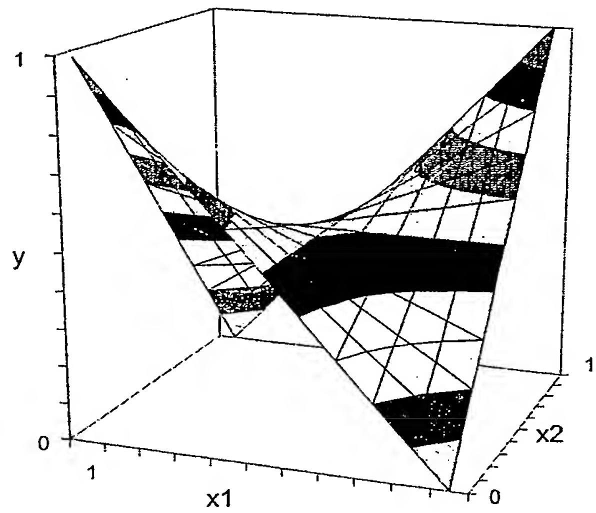

1
y
0
1
x1
1
x2
0

FIG. 4

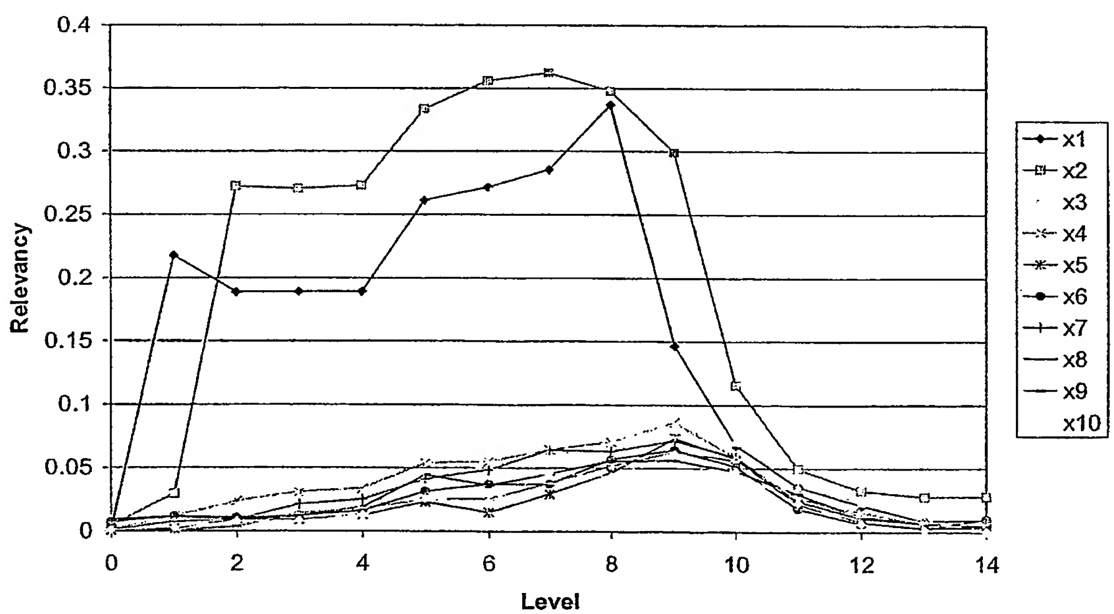
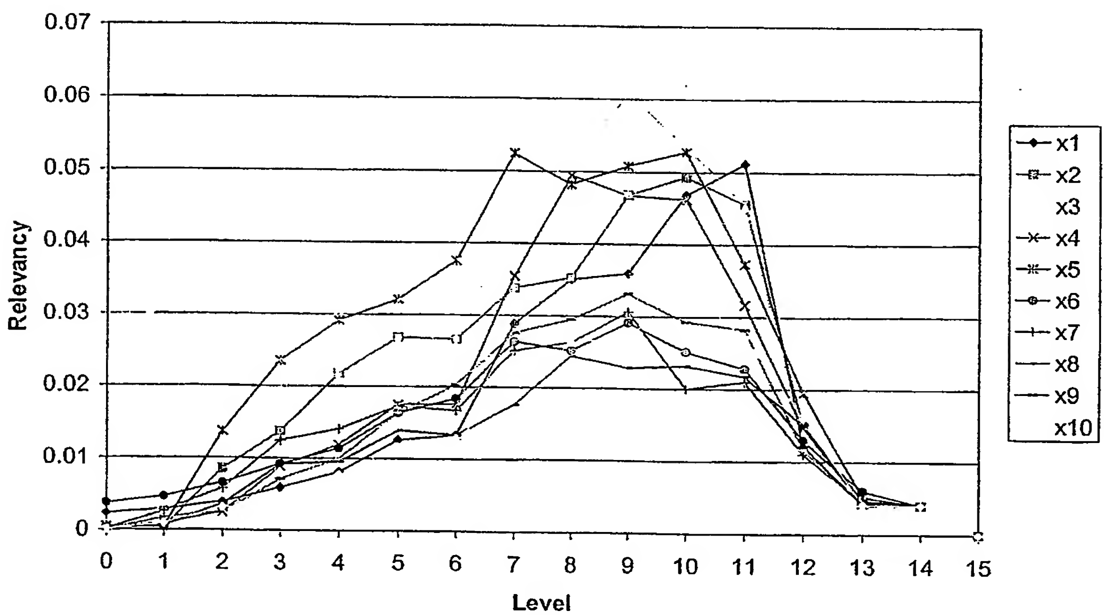

0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Relevancy
0
2
4
6
8
10
12
14
Level
x1
x2
x3
x4
x5
x6
x7
x8
x9
x10

FIG. 5

FIG. 6

Training algorithm	2 inputs				10 inputs			
	time	DOF	MSE	ANOVA R^2	time	DOF	MSE	ANOVA R^2
RBF	1	15	0.00008	99.17	2	209	0.00015	98.52
BP	25	12	0.00147	85.38	39	36	0.00214	78.62

FIG. 7



INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/021511

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06K9/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KRUSINSKA E: "TWO STEP SEMI-OPTIMAL BRANCH AND BOUND ALGORITHM FOR FEATURE SELECTION IN MIXED VARIABLE DISCRIMINATION" PATTERN RECOGNITION, ELSEVIER, KIDLINGTON, GB, vol. 22, no. 4, 1989, pages 455-459, XP000046518 ISSN: 0031-3203 Section 2. Criterion for selection, Section 3. Branch and bound method	1,4-15
Y	the whole document	2,3, 16-28

-/--

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 November 2004	11/11/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Grigorescu, C

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/021511

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	W.-S. HWANG, J. WENG: "Hierarchical discriminant regression" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 22, no. 11, November 2000 (2000-11), XP002302472 Section 2.2. Distance in discriminating space	2,3, 16-28
X	JAIN A ET AL: "Feature selection: evaluation, application, and small sample performance" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE INC. NEW YORK, US, vol. 19, no. 2, February 1997 (1997-02), pages 153-158, XP002294891 ISSN: 0162-8828	1
A	the whole document	2-28
A	FRANK I E: "Modern nonlinear regression methods" 1995, CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS, ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM, NL, PAGE(S) 1-19 , XP004037947 ISSN: 0169-7439 the whole document	1-28
A	NARENDRA P M ET AL: "A BRANCH AND BOUND ALGORITHM FOR FEATURE SUBSET SELECTION" IEEE TRANSACTIONS ON COMPUTERS, IEEE INC. NEW YORK, US, vol. C-26, no. 9, 1 September 1977 (1977-09-01), pages 917-922, XP000647423 ISSN: 0018-9340 the whole document	1-28